Computer Networks xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Optimal online control for sleep mode in green base stations

Richard Combes [a], Salah Eddine Elayoubi [b,*], Arshad Ali [b], Louai Saker [b], Tijani Chahed [c]

[a] Supelec, 3, rue Joliot-Curie, 91192 Gif-sur-Yvette, France
[b] Orange Labs, 38-40 rue du Gnral Leclerc, 92130 Issy-les-Moulineaux, France
[c] Telecom SudParis, 9 rue Charles Fourier, 91000 Evry, France

ARTICLE INFO

Article history:
Received 19 March 2014
Received in revised form 22 September 2014
Accepted 1 October 2014
Available online xxxx

Keywords:
Mobile networks
Sleep mode
Optimal control

ABSTRACT

In this paper, we investigate network sleep mode schemes for reducing energy consumption of radio access networks. We first propose, using Markov Decision Processes (MDPs), an optimal controller that associates to each traffic an activation/deactivation policy that maximizes a multiple objective function of the Quality of Service (QoS) and the energy consumption. We focus on a practical implementation issue, namely the ping pong effect resulting in unnecessary ON/OFF oscillations, that may affect the stability of the system. We illustrate our results numerically using theoretical models of the radio access network, and apply the developed mechanisms on a large-scale network simulator. Knowing that an offline optimization is not suitable for a large-scale network nor does it fit all traffic configurations, we propose, using an online controller that derives dynamically the optimal policy based on the dynamics of users in the cell. The design of our online controller is based on a simple ϵ-greedy algorithm and learns the optimal threshold policy for activation/deactivation of network resources.

1. Introduction

Energy efficiency has become a critical issue in future wireless Internet due to the huge amount of energy needed in order to create, maintain and co-operate mobile networks: 2G, 3G/3G+ as well as future 3G Long Term Evolution (LTE) and LTE-Advanced systems [1]. This is called green radio where energy consumption is a critical constraint when designing and operating the network [2].

Classically, the energy consumption of the base stations has not been considered as a constraint when designing Joint Radio Resource Management (JRRM) schemes. The aim was always to ensure higher spectral efficiencies and better QoS (e.g. [3,4]). A large interest has been dedicated to energy savings at the user equipment; the aim being to preserve the user equipment's battery by reducing the amount of energy that is not useful for transmitting information. Consequently, many sleep mode schemes for user equipment have been proposed in the literature. For instance, in [5], optimal sleep mode parameters have been derived depending on the traffic pattern for mobile WiMAX devices. Authors in [6] assessed the performance of discontinuous transmission schemes for UMTS. Particular attention has been set on the sleep mode in indoor access points [7,8], where intelligent wake up mechanisms have been proposed.

However, when studying the energy footprint of mobile networks, it is observed that macro base stations are the most energy consuming nodes (around 80% of the overall mobile network consumption [9]) and must thus be at the heart of any green radio scheme. An important set of works on green radio has then been dedicated to the reduction of the transmitted power of the base stations; the idea is to find the optimal transmission power that ensures coverage and capacity (see for instance [10,11]). This approach is essential for reducing the exposure of

* Corresponding author.

http://dx.doi.org/10.1016/j.comnet.2014.10.031
1389-1286/© 2014 Elsevier B.V. All rights reserved.

persons to electromagnetic radiations. However, alone, these schemes are not sufficient to reduce the energy consumption of wireless networks since a large amount of energy is consumed even for low output power. This is due to the load-independent components of the energy consumption and the presence of pilot channels that make low load resources totally inefficient in terms of energy. This is also the reason that makes energy-aware load balancing techniques not so efficient (an average reduction of the energy consumption of 5% has been observed in [12]).

However, if a modular architecture allows switching off the network resources that are not necessary to guarantee the target QoS for the offered traffic, the gains may be very large. *Network* sleep mode is then crucial if the aim is to reduce significantly energy consumption of the network. In this case, when a network resource, for instance frequency carrier, is not needed to ensure QoS, it is turned off to reduce global energy consumption. This opportunity of implementing sleep mode in macro base stations has been addressed in [9,13], where it has been shown that continuous coverage can be ensured in low traffic scenarios with a small number of active base stations. This paper fills a gap in the performance evaluation of network sleep mode while ensuring user-perceived QoS. We consider local implementation of sleep mode mechanisms that have the advantage of being easier to implement and more reactive (there is no need to introduce a new entity in the network (a controller), and create interfaces between the controller and the stations). Local implementation enables faster reaction to traffic variations since feedback of measurements to a centralized controller takes a non-negligible amount of time. Furthermore, we only consider in this work sleep mode of resources within the base station, without having to switch off complete cell sites so that the coverage is preserved (we ensure that at least one carrier or system is active at any time).

Our contribution: We propose sleep mode policies that associate, to each system state, the optimal action consisting in activating or deactivating a subset of the resources. We first study the optimal controller that proposes the best policy corresponding to the traffic and system settings, using a Continuous Time Markov Decision Process (CTMDP) [16]. We focus on the ping pong effect that may happen due to consecutive ON and OFF commands, and show that a hysteresis time introduced in the decision mechanism can efficiently cope with this issue. Our numerical results show that the optimal policy is a threshold policy on the number of active users in the cells. We thus propose an online version of our controller, using a simple ϵ-greedy algorithm [14] that explores the different threshold policies until finding the optimal one, in an efficient way. Our numerical results show that the online controller converges fast to the optimal policy and does not need any a priori knowledge about traffic or radio conditions.

Related work: Previous results such as [10,11] demonstrate how to choose the minimal amount of power to be transmitted by base stations in order to ensure coverage and a minimal data rate uniformly over the network. In [22], a wireless network with multiple base stations and users is considered. Users are considered fixed, and the authors formulate the problem of base station activation as a combinatorial optimization problem. Contrary to the present work, the proposed solution is *computationally heavy*. Previous papers such as [23,24] consider a single user and use a Markov chain technique to evaluate the energy savings due to the sleep mode mechanism of a single user terminal. In particular [24] allows taking into account correlated packet arrivals. [25] considers a similar setting (one user and one station), and show how to derive the optimal sleep policy numerically by formalizing the problem as a Semi-MDP. In contrast to [23,24] we are more concerned with base station sleep mode (rather than user equipment sleep mode), and furthermore we consider a multi-cellular network with multiple users. In all of the work above, the fact that users arrive and depart dynamically (i.e. the *flow-level dynamics*) is not taken into account, contrary to the present work. Furthermore, we believe that the idea of adjusting sleep mode parameters based on learning (estimating the initially unknown traffic intensity) was not investigated in the works cited above. Our previous work [15] considered flow level dynamics while designing sleep mode mechanisms, but the optimal policy was derived using an exhaustive search of all possible policies. In this paper, we propose a framework that derives the optimal policy using MDP theory, while the work in [15] was based on a derivation of the optimal policy by an exhaustive search among all possible policies. We then implement our controller on a large scale network simulator [21] and show how the optimal policy varies from one base station to another. We finally propose an online controller, based on an ϵ-greedy policy, and show how it converges to the optimal solution.

The remainder of this paper is organized as follows. The system model is described in Section 2. In Section 3, we propose an optimal control mechanism for the network sleep mode and show how it can be applied to actual and future mobile networks. Section 3 also deals with the ping pong effect and proposes a solution to take it into account in the decision. Section 4 analyzes the performance of the proposed scheme, using theoretical models of the radio interface, and also presents results issued from a large-scale network simulator and shows how the optimal scheme may vary between cells. Section 5 proposes a simple, yet efficient, online controller based on the ϵ-greedy policy that converges towards the optimal policy. Section 6 concludes the paper.

2. System model

2.1. Definition of sleep mode policies

We consider a cellular network as the one depicted in Fig. 1 and focus on the downlink of a particular sector equipped with a pool of R resources, for instance a number of transmitters (TRX) for GSM, carriers for 3G, HSDPA and LTE, or even small cells or relays in dense deployed networks. Let $\mathbf{s}_t$ be the process describing the evolution of the system state and $\mathcal{S}$ be the state space. This state will be defined in the next section for 2G, 3G and 4G systems.

Please cite this article in press as: R. Combes et al., Optimal online control for sleep mode in green base stations, Comput. Netw. (2015), http://dx.doi.org/10.1016/j.comnet.2014.10.031

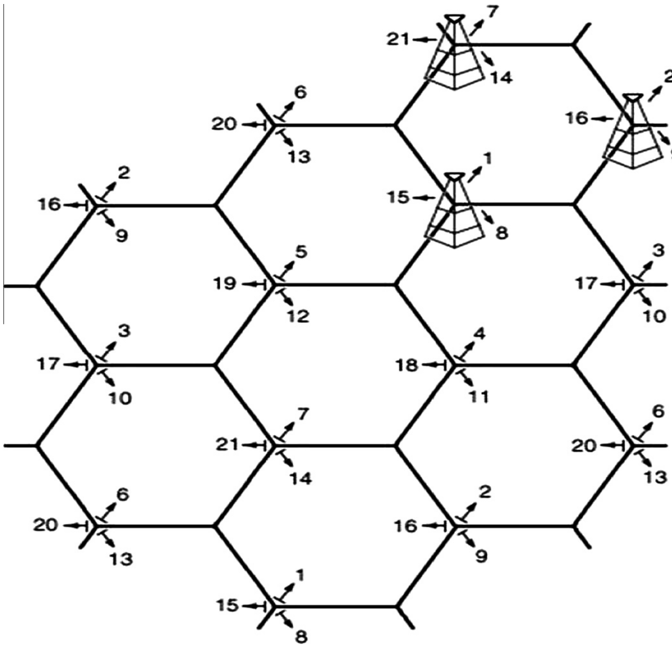

Fig. 1. Cellular layout. Each macro base stations is equipped with three directional antennas, creating three sectors.

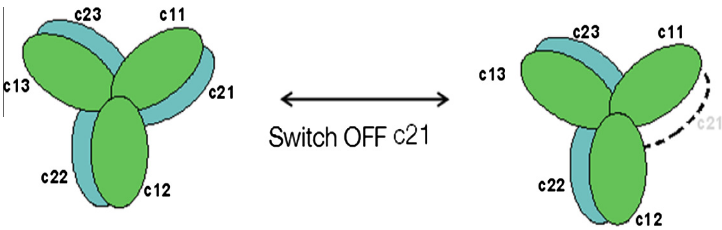

Fig. 2. Example of sleep mode in a base station where each sector is equipped with two power amplifiers, each one serving a cell (e.g. 3G carriers). One of the cells is switched off in the right part of the figure.

The Radio Resource Management (RRM) scheme distributes the users between the R resources, and generates a certain load $\rho_j(\mathbf{s}|R)$ for resource $j \in \{1, \ldots, R\}$ in state $\mathbf{s}$. The energy consumption of the base station depends on these loads and is given by [12]:

$$E(\mathbf{s}) = P_{cst} + \sum_{j=1}^{R} \left(P_{TRX} + \rho_j(\mathbf{s}|R) \frac{P_{max}}{c_{DC}} \right), \tag{1}$$

where c_{DC} is the DC to RF conversion factor, P_{max} is the radio frequency output power of the power amplifier, P_{TRX} is the fixed power consumption of the radio module (transceiver) and P_{cst} is a fixed power consumption due to transport and processing units.

When sleep mode is implemented, we define frontiers F_r within the state space; each time these frontiers are crossed, a resource is activated (or deactivated). $\mathcal{S}$ is thus split between subspaces $\mathcal{S}_r$, each corresponding to a number of activated resources r. The sleep mode scheme corresponds to switching ON/OFF a resource each time the frontier between the different subspaces is crossed in a direction or another. Here, we assume that, when traffic increases and a frontier is crossed upwards, a resource is immediately set back into service. Fig. 2 illustrates this sleep mode mechanism with an example of a tri-sectored site where each sector has two power amplifiers serving each a carrier; one of the power amplifiers (that serving carrier 2 in sector 1) is switched off in the right part of the figure.

In order to study the performance of sleep mode, we modify the state space by introducing the number of activated resources. Let the new state space be denoted by $\tilde{\mathcal{S}}$, composed of the new states $\tilde{\mathbf{s}}$:

$$\tilde{\mathcal{S}} = \{\tilde{\mathbf{s}} = (\mathbf{s}, r); \mathbf{s} \in \mathcal{S}, r \in \{1, \ldots, R\}\}. \tag{2}$$

Our aim is to find the optimal sleep mode that minimizes energy consumption while preserving QoS. Let us define the resource activation policy $\mathbf{P}$, associating an action $P(\mathbf{s}, e)$ to event e (arrivals or departures of calls).[1] The

[1] The degenerate policy $\mathbf{P} = 0$ corresponds to the non constrained process where all resources are active.

action can be activating a resource (+1), deactivating it (−1), or doing nothing (0). Define $\mathcal{P}$ as the set of all possible policies. Let $\mathbf{Q}$ be the transition matrix of the initial, non constrained, process, with $q(\mathbf{s}, \mathbf{s}')$ the transition rate between states $\mathbf{s}$ and $\mathbf{s}'$ in $\mathcal{S}$. The transition matrix $\tilde{\mathbf{Q}}(\mathbf{P})$ of the new process, knowing the policy $\mathbf{P}$, can be obtained based on $\mathbf{Q}$:

$$\tilde{q}((\mathbf{s}, r), (\mathbf{s}', r + a)) = q(\mathbf{s}, \mathbf{s}') I_{P(\mathbf{s}, e(\mathbf{s}, \mathbf{s}')) = a}, \tag{3}$$

with a the action (0, +1 or −1), $e(\mathbf{s}, \mathbf{s}')$ is the event causing transition from state $\mathbf{s}$ to state $\mathbf{s}'$ in the initial process, and I_x is the indicator function whose value is equal to 1 if event x is true, and to 0 otherwise.

Note that, by construction, we define the policy as function on the actual state. The decision of switching ON/OFF a resource is thus taken by observing the actual state only. If the initial process $(\mathbf{s}_t, \mathbf{Q})$ is Markovian, the modified process $(\tilde{\mathbf{s}}_t, \tilde{\mathbf{Q}}(\mathbf{P}))$ is Markovian. In this case, let $\tilde{\Pi}(\mathbf{P})$ be the vector of steady-state probabilities $\tilde{\pi}(\tilde{\mathbf{s}}|\mathbf{P}), \tilde{\mathbf{s}} \in \tilde{\mathcal{S}}$.

Knowing these steady-state probabilities, the energy consumption of the network is calculated by:

$$\overline{E}(\mathbf{P}) = \sum_{\tilde{\mathbf{s}} \in \tilde{\mathcal{S}}} E(\tilde{\mathbf{s}}) \tilde{\pi}(\tilde{\mathbf{s}}|\mathbf{P}), \tag{4}$$

with $E(\tilde{\mathbf{s}})$ the energy consumption knowing state $\tilde{\mathbf{s}}$ (number of users and number of activated resources), obtained as in Eq. (1), replacing R by r.

On the other hand, the sleep mode scheme may have an impact on the QoS. For instance, if we decide to deactivate some resources for a set of non-empty states, the throughput perceived by the ongoing users will decrease and the overall capacity degrades compared to the baseline case. Let $f(\mathbf{P})$ be the average QoS under policy $\mathbf{P}$. In the following we will show how to derive this QoS for practical cases.

The optimization problem consists in finding the policy $\mathbf{P}^*$ that verifies:

$$\mathbf{P}^* = \arg\min_{\mathbf{P} \in \mathcal{P}} E(\mathbf{P}) \tag{5}$$

subject to:

$$f(\mathbf{P}) \geqslant QoS_{min}, \tag{6}$$

where QoS_{min} is the minimal acceptable QoS.

2.2. Application to real time traffic

In circuit switched mobile networks, such as GSM and UMTS, and in packet switched networks carrying real-time traffic (e.g. streaming on Beyond 3G networks), a certain set of radio conditions can be defined, leading to a definition of the system state:

- **In GSM networks**, frequency reuse, by eliminating inter-cell interference, harmonizes the radio conditions in the cell, leading to a constant consumption of radio resources, independent from the position in the cell. Each cell has a number of transmitters (say r) each operating on a specific frequency. A transmitter is then allocated to users following a Time Division Multiple Access (TDMA) manner; time is divided into N_{trx} time slots, among which a certain number N_{sig} is reserved for signaling. The state of the cell is then described by the number of active users, denoted by n ($\mathbf{s} = (n)$), and the capacity equation of the cell when r transmitters are active is given by:

$$n \leqslant r(N_{trx} - N_{sig}). \tag{7}$$

If we suppose perfect load balancing between active transmitters, a state $\tilde{\mathbf{s}} = (n, r)$ generates a cell load (percentage of total base station power) equal to:

$$\rho(\tilde{\mathbf{s}}) = \frac{r.N_{sig} + n}{r.N_{trx}}. \tag{8}$$

- **In UMTS networks**, cell edge users consume more power resources than cell center ones, leading to different classes of radio conditions ($c \in [1, C]$). The state of the system is then described by $\mathbf{s} = (n_1, \ldots, n_C)$, with n_c the number of users subject to radio condition of type c. The capacity equation of a 3G carrier is given by [17]:

$$\sum_{c=1}^{C} b_c n_c \leqslant r \cdot (P_{max} - P_{com}), \tag{9}$$

where b_c is the power consumed by a user of radio condition c, calculated as in [17]. P_{max} is the total power on a carrier and P_{com} the power of signaling channels. The cell load is then calculated by:

$$\rho(\tilde{\mathbf{s}}) = \frac{\sum_{c=1}^{C} b_c n_c + r.P_{com}}{r.P_{max}}. \tag{10}$$

- **In Beyond 3G networks** (HSDPA and 3G LTE), the heterogeneous radio conditions are reflected by different throughputs D_c that can be achieved on a carrier at different positions in the cell [18]. The state of the system, when carrying real time traffic, is then also characterized by the number of users in the different positions $\mathbf{s} = (n_1, \ldots, n_C)$, and the capacity equation when there are r active carriers is given by:

$$\sum_{c=1}^{C} \frac{T}{D_c} n_c \leqslant r, \tag{11}$$

where T is the target throughput. The cell load is thus:

$$\rho(\tilde{\mathbf{s}}) = \sum_{c=1}^{C} \frac{T}{r D_c} n_c. \tag{12}$$

Note that this load equation supposes carrier aggregation feature, where several carriers can be accessed simultaneously by each user, offering a multiple of the throughput that can be achieved on a single carrier. Such carrier aggregation features have been standardized in HSDPA+ and LTE-Advanced networks.

In all these networks, in the absence of sleep mode, the transitions between states are due to arrivals and departures of users. The elements of the $\mathbf{Q}$ matrix, indicating the transitions between states $\mathbf{s} = (n_1, \ldots, n_c)$ and $\mathbf{s}' = (n'_1, \ldots, n'_c)$, are given by:

$$q(\mathbf{s}, \mathbf{s}') = \begin{cases} \lambda_c, & n'_c = n_c + 1 \\ n_c \mu, & n'_c = n_c - 1 \end{cases} \tag{13}$$

where λ_c is the arrival rate of users of radio condition c, and μ the service rate of calls.

After modifying the chain by introducing the sleep mode policy $\mathbf{P}$, the actions are introduced in the new transition matrix $\tilde{\mathbf{Q}}(\mathbf{P})$, as indicated in Eq. (3). Note that, as we assume that resources can be instantaneously activated, there is no interest in activating a new resource unless the new arrival will make the load in Eqs. (8), (10) or (12) exceed 1. However, it would be interesting to accept some QoS degradation by retarding resource activation, if the gain in energy is large and constraint (6) is not violated. This latter is now a constraint on blocking rates.

Remark 1. In a network without sleep mode carrying streaming traffic, multi-Erlang methods can be used to evaluate the performance, and the steady-state probabilities have the product form:

$$\pi(\mathbf{s}) = \frac{1}{\sum_{\mathbf{s}' \in \mathbf{S}} \Pi_{c=1}^{C} \frac{(\lambda_c/\mu)^{n'_c}}{n'_c!}} \Pi_{c=1}^{C} \frac{(\lambda_c/\mu)^{n_c}}{n_c!}. \tag{14}$$

When a simple sleep mode is applied, where exactly the needed number of carriers is used at each moment, this product form still holds and the blocking rate is still calculated using multi-Erlang; only the energy consumption is reduced.

2.3. Application to elastic traffic

When elastic traffic is considered in Beyond 3G networks, the heterogeneity in radio conditions leads to a larger service time for cell edge users. The service rate becomes state-dependent, and the transitions between states $\mathbf{s} = (n_1, \ldots, n_c)$ and $\mathbf{s}' = (n'_1, \ldots, n'_c)$, are given by:

$$q(\mathbf{s}, \mathbf{s}') = \begin{cases} \lambda_c, & n'_c = n_c + 1 \\ \frac{n_c D_c}{F \sum_{i=1}^{C} n_i}, & n'_c = n_c - 1 \end{cases} \tag{15}$$

where F is the average file size.

Remark 2. In a network without sleep mode carrying elastic traffic, the cell can be modeled as a Processor Sharing queue, and its evolution described by the overall number of users in a cell $n = \sum_{c=1}^{C} n_c$. The solution of (19) has the simple form [18] $\pi(n) = (1-\rho)\rho^n$, with the cell load calculated by $\rho = \lambda F/(RD), \lambda$ being the overall arrival rate and $D = \left(\sum_{c=1}^{C} \frac{\lambda_c}{\lambda D_c}\right)^{-1}$, the harmonic mean of the throughput. If a target throughput T is sought, the probability, for users of radio condition c, to achieve this throughput is given by:

$$QoS_c = \sum_{m=1}^{K_c} \pi(m) = 1 - (K_c + 1)\rho^{K_c} + K_c \rho^{K_c+1}, \tag{16}$$

where $K_c = \lfloor \frac{D_c}{T} \rfloor$ is the maximal allowable number of users in the cell such that the throughput of a user in position c is acceptable. The overall QoS is given by:

$$\bar{QoS} = \sum_{c=1}^{C} \frac{\lambda_c}{\lambda} QoS_c. \tag{17}$$

Note that, even when a sleep mode is applied, but with a simple policy that activates the carriers depending on the overall number of users n, a Generalized processor sharing describes the evolution of the system:

$$\tilde{\pi}(n|\mathbf{P}) = \tilde{\pi}(0|\mathbf{P}) \frac{(\lambda F/D)^n}{\Pi_{m=1}^{n} r(m|\mathbf{P})}, \tag{18}$$

$r(n|\mathbf{P})$ being the number of active resources in state n, if we apply policy $\mathbf{P}$.

3. Optimal control of network sleep mode

The analysis of the previous section can be used as a framework for optimizing the sleep mode policy. Indeed, the steady-state probabilities, for a given policy, can be obtained by solving:

$$\tilde{\Pi}(\mathbf{P}).\tilde{\mathbf{Q}}(\mathbf{P}) = 0 \tag{19}$$

with a normalization to 1 of these probabilities.

For a given traffic intensity and known radio conditions, an exhaustive search over all possible policies allows finding the optimal one. However, a controller where the network adjusts its policy function of the actual situation is more suitable for large-scale heterogeneous networks. This controller observes the current state $\tilde{\mathbf{s}}$ of the network, and, at each event (arrival or departure) takes a decision $a(\tilde{\mathbf{s}})$ that moves the system to a new state, and incurs a cost. This corresponds to a Continuous Time Markov Decision Process (CTMDP), with the following formal representation [19]:

$$\left(\tilde{\mathbf{S}}, \mathbf{A}, (\mathbf{A}_{\tilde{\mathbf{s}}}, \tilde{\mathbf{s}} \in \tilde{\mathbf{S}}), \tilde{q}(\tilde{\mathbf{s}}'|\tilde{\mathbf{s}}, a), g(\tilde{\mathbf{s}}, a)\right), \tag{20}$$

where $\tilde{\mathbf{s}} = (\mathbf{s}, r)$ is the state and $\mathbf{A} = \{0, 1, -1\}$ is the set of possible actions corresponding to "doing nothing", "adding a resource" and "switching OFF a resource". The set of possible actions reduces to a subset $\mathbf{A}_{\tilde{\mathbf{s}}}$ for particular states (e.g. adding a resource is impossible when $r = R$). $\tilde{q}(\tilde{\mathbf{s}}'|\tilde{\mathbf{s}}, a)$ are the transition rates knowing the decision a. Finally, $g(\tilde{\mathbf{s}}, a)$ is the cost function, subsequent to action a in state $\tilde{\mathbf{s}}$; this cost function has to be an increasing function of energy and a decreasing function of QoS.

If the initial process $(\mathbf{s}_t, \mathbf{Q})$ is Markovian and the elements of matrix $\mathbf{Q}$ are bounded, then there exists an optimal Markovian deterministic policy for the corresponding CTMDP. This can be found by uniformization and discretization of the initial process as follows [16]:

- When all the transition rates in matrix $\mathbf{Q}$ are bounded, the sojourn times in all states are exponential with bounded parameters $q(\mathbf{s}, \mathbf{s})$. $\sup_{\mathbf{s} \in \mathbf{S}}(\mathbf{s}, \mathbf{s})$ thus exists, and we can find a constant v verifying:

$$\sup_{\mathbf{s} \in \mathbf{S}} [1 - p(\mathbf{s}, \mathbf{s})] q(\mathbf{s}, \mathbf{s}) \leqslant v < \infty, \tag{21}$$

  with $p(\mathbf{s}, \mathbf{s})$ the probabilities of staying in the same state after the next event.
- We can thus define an equivalent, uniformized process with state-independent exponential sojourn times of parameter v, and transition probabilities:

$$\tilde{p}(\mathbf{s}, \mathbf{s}') = \begin{cases} 1 - \frac{[1 - p(\mathbf{s}, \mathbf{s})] q(\mathbf{s}, \mathbf{s})}{v}, & \mathbf{s}' = \mathbf{s}, \\ \frac{p(\mathbf{s}, \mathbf{s}') q(\mathbf{s}, \mathbf{s})}{v}, & \mathbf{s}' \neq \mathbf{s}. \end{cases} \tag{22}$$

Please cite this article in press as: R. Combes et al., Optimal online control for sleep mode in green base stations, Comput. Netw. (2015), http://dx.doi.org/10.1016/j.comnet.2014.10.031

These two processes, the original and the uniformized one, are equal in distribution. As the state space of this MDP is discrete and the set of possible actions $\mathbf{A}$ is finite (there is a limited number of resources R), there exists an optimal Markovian deterministic policy ([16], Theorem 6.2.10), for the uniformized process, that will be also optimal for the CTMDP.

Note: Resource activation cannot be assumed to be instantaneous and activation delays have to be taken into account. However since the proposed mechanism is assumed to be implemented directly into the base station, one can assume that the activation delay is small (in the order of 100 ms) as there is no need for signaling between neighboring sites and the signal processing units of the base station are active (at least one resource is kept active to preserve coverage). As the time scale on which flows arrive and depart is in the order of seconds, the activation appears instantaneous at that time scale. If for whatever reason, for instance sleep mode of complete cell sites, the activation delay were to be significant, one can modify the mathematical formalism by a straightforward modification of the technique described in [26]. Specifically, the system becomes an MDP with delays which can be transformed into an MDP without delays by increasing the dimension of the state.

3.1. Application to beyond 3G networks

We now illustrate the solution of this CTMDP in the case of Beyond 3G networks, carrying elastic traffic, where the decision is based on the number of users in the cell. The case of real time traffic can be analyzed in a similar manner, using transition matrix $\mathbf{Q}$ of Section 2.2.

From state (n,r), the transitions are given by:

- Upon arrivals, the transition occurs with rate λ to the state with $n+1$ users, but with a possibly larger number of active resources.
- Upon departures, the transition occurs with rate $r.\mu$ to the state with $n-1$ users, but with a possibly smaller number of active resources. $\mu = D/F$ is the service rate if there is only one activated carrier.

The transition rates are thus equal to:

$$\tilde{q}(n',r'|n,r,a) = \begin{cases} \lambda, & n'=n+1, r'=r+a, \\ r\mu, & n'=n-1, r'=r+a, \\ 0, & \textit{otherwise}. \end{cases} \quad (23)$$

As of the cost function, it is related to the energy consumption and the QoS. Let us consider for instance the following cost function:

$$g(n,r,a) = \sum_{(n',r')\in\tilde{\mathcal{S}}} \{ I_{\tilde{q}(n',r'|n,r,a)\neq 0} \times [\gamma \bar{E}(n', r+a) \\ +(1-\gamma)(1-f(n',r+a))]\}, \quad (24)$$

where $\gamma \in [0,1]$ is a coefficient that prioritizes between energy consumption and QoS. $f(.)$ is a QoS measures whose value is 1 if the QoS is best and to 1 if it is unacceptable (i.e. a satisfaction function). $\bar{E}$ is a normalized energy consumption (1 if maximal). Note that, in this expression, only the term that corresponds to the transition subsequent to action a is not equal to zero. The energy consumption is derived from Eq. (1):

$$E(n,r) = \begin{cases} P_{cst} + r(P_{TRX} + \frac{P_{com}}{c}), & n=0, \\ P_{cst} + r(P_{TRX} + \frac{P_{max}}{c}), & n>0. \end{cases} \quad (25)$$

The QoS is function of the throughput achieved by users. If a target throughput T is fixed, an example QoS function is:

$$f(n,r) = \exp\left(-\frac{\log(2)T}{rD/n}\right). \quad (26)$$

This function ensures a satisfaction of 1/2 when the throughput $\frac{rD}{n}$ is equal to the target throughput T. It tends to 1 when the throughput is large and to zero when the throughput is too low. This objective function was proposed by Enderlé and Lagrange in [20] to model the customer satisfaction when elastic traffic is considered.

Now that we have the definition of the CTMDP, we can proceed to its unformization. The maximal departure rate being equal to $R\mu$ (all the resources are active), $\nu = \lambda + R\mu$ verifies (21) and can be taken as a basis for the uniformization. The transition rates from state (n,r) in the uniformized MDP are then given by:

$$\hat{q}(n',r'|n,r,a) = \begin{cases} \lambda/\nu, & n'=n+1, r'=r+a, \\ r\mu/\nu, & n'=n-1, r'=r+a, \\ \frac{\nu-r\mu-\lambda}{\nu}, & n'=n, r'=r, \\ 0, & \textit{otherwise}. \end{cases} \quad (27)$$

If the discounted reward method, with discount rate α, is used, the cost function after uniformization has to be also modified as follows ([16], Theorem 11.5.2):

$$\hat{g}(n,r,a) = g(n,r,a) \sum_{(n',r')} \frac{\alpha + \tilde{q}(n',r'|n,r,a)}{\alpha+\nu} I_{\tilde{q}(n',r'|n,r,a)\neq 0}. \quad (28)$$

3.2. Coping with ping pong effect

In the previous section, we showed how to derive optimal controllers for sleep mode in mobile networks. When traffic in the cell increases/decreases, some resources are to be activated/deactivated for reducing energy consumption. However, a ping-pong effect may occur, where the resource is activated/deactivated several times, leading to a large signaling overhead. This effect is accentuated when there is a non negligible activation time, as resources may be deactivated due to temporary load decrease, and cannot be activated rapidly to respond to traffic demand.

We will study in this section mechanisms that minimize the ping pong effect. In order to avoid oscillations at the frontier, we propose a hysteresis time T_{off} before deactivating a resource when the frontier $\check{F}_{r'}$ is downcrossed.

3.2.1. Quantifying the ping pong effect

The first step towards reducing the ping pong effect is to quantify it. Let $\mathcal{F}$ be the set of couples of states $((\mathbf{s},r),(\mathbf{s}',r+1)), \mathbf{s},\mathbf{s}' \in \mathcal{S}, r \in \{1,\ldots,R-1\}$.

Proposition 3.1. *A good measure of the ping pong effect is given by:*

$$Pp = \sum_{(\tilde{\mathbf{s}},\tilde{\mathbf{s}}')\in\mathcal{F}} \pi(\tilde{\mathbf{s}})q(\tilde{\mathbf{s}},\tilde{\mathbf{s}}'), \tag{29}$$

where $\pi(\tilde{\mathbf{s}})$ is the steady-state probability of state $\tilde{\mathbf{s}}$ and $q(\tilde{\mathbf{s}},\tilde{\mathbf{s}}')$ the transition rate between states $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{s}}'$.

Eq. (29) gives in fact the rate of activating a resource, when the system is in steady-state.

3.2.2. Optimizing hysteresis time

In order to reduce the ping pong effect, we propose a hysteresis mechanism: When the frontier corresponding to deactivating a resource is down-crossed, an exponential timer is activated, and the resource is switched off only if the load is still low at its expiration. This defines a new branch of the chain characterized by a variable o whose value "1" indicates a deactivation timer. The transitions from the main branch ($o = 0$) to the switching off branch ($o = 1$) occurs when the frontier F_r is down-crossed. The system state evolves then within this chain until this frontier is crossed again (the deactivation decision is canceled), or the timer expires. In this latter case, there is a transition between state $(\mathbf{s}, r, v = 0, o = 1)$ to state $(\mathbf{s}, r-1, v = 0, o = 0)$, with rate $\frac{1}{T_{off}}$, where T_{off} is the average timer duration.

Note that the hysteresis mechanism is to be associated with activation anticipation. The optimal policy (obtained by (5), (6)) will have thus two components: the action to be taken at each state ($a \in \{0, +1, -1\}$) and the duration of the deactivation timer (T_{off}). The switch off timer is to be integrated within the original process before uniformization (i.e. the action $a = -1$ causes a transition to a state with $o = 1$). Indeed, the hysteresis mechanism is a control mechanism performed by the user in order to enhance the performance, in opposition with the activation time that is an external delay imposed to the system. To illustrate this, we revisit the elastic traffic case of Section 3.1. The uniformization rate becomes:

$$v = \lambda + R\mu + \frac{1}{T_{off}} \tag{30}$$

and the cost function has to include a component related to the ping pong effect:

$$g(n,r,o,a) = \sum_{(n',r')\in\tilde{\mathcal{S}}} \Big\{ I_{\tilde{q}(n',r',o'|n,r,o,a)\neq 0} \times [\gamma_e \overline{E}(n', r+a) + \gamma_q(1 - f(n', r+a)) + \gamma_p(r'-r)^2] \Big\}, \tag{31}$$

where γ_e, γ_q and γ_p are three positive constants less than one verifying $\gamma_e + \gamma_q + \gamma_p = 1$, and $(r'-r)^2)$ is equal to 1 if an effective switch ON/OFF operation has occurred.

4. Numerical analysis and simulations

In this section, we illustrate the performance of the sleep mode schemes based on theoretical models of the radio interface, developed in [17] for UMTS and [18] for HSDPA. We show the performance of the proposed schemes considering a sleep mode.

4.1. Numerical analysis: exhaustive search of the optimal policy

We first consider a UMTS cell carrying voice traffic and equipped by two frequency carriers of 5 MHz each. In order to illustrate the performance of the sleep mode mechanism, we make use of an exhaustive search over all possible policies. This consists in obtaining the steady-state probabilities for each policy (Eq. (19)), computing the corresponding energy consumption (Eq. (4)) and selecting the policy that ensures the target QoS with the lowest energy (Eq. (5)). We plot in Fig. 3 the energy consumption when both carriers are always activated and compare it to a perfect sleep mode scheme where the exact number of carriers is activated at each state. Here, there is no degradation of the QoS as blocking occurs, in both cases, only when more than 2 carriers are needed. We can observe that a large energy consumption gains can be achieved for low traffic figures, as one carrier is needed almost all the time. However, when the traffic increases, the sleep mode scheme is no more efficient as all the resources are needed in order to ensure the QoS.

In the remainder of this paper, we consider data traffic on HSDPA networks. We consider a cell equipped with two carriers, each one of maximal capacity equal to 3 Mbps (harmonic average of the single-user throughput over the cell surface). We suppose that the carrier aggregation feature is implemented, so that a user can profit from the throughput offered by all the activated carriers. For the FTP-like traffic, we consider that the operator wants to offer a throughput of 400 Kbps to a large proportion of its clients (95%).

In this case, the optimal sleep mode policy is not trivial, as we consider a soft perception of the QoS based on the satisfaction of users. The optimal policy is thus to be derived for each traffic scenario. We vary the cell offered traffic and plot the energy consumption corresponding to the optimal strategy of the controller (Fig. 4). A similar behavior as for voice traffic can be observed: Large gains for low to medium traffic, and small gains when the cell becomes saturated.

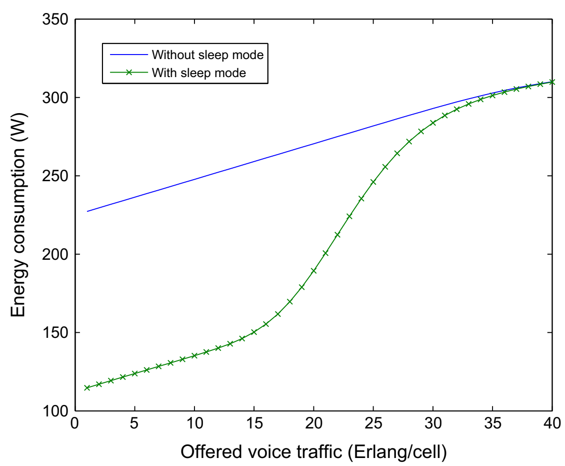

Fig. 3. Energy consumption of a 3G base station for different offered voice traffic.

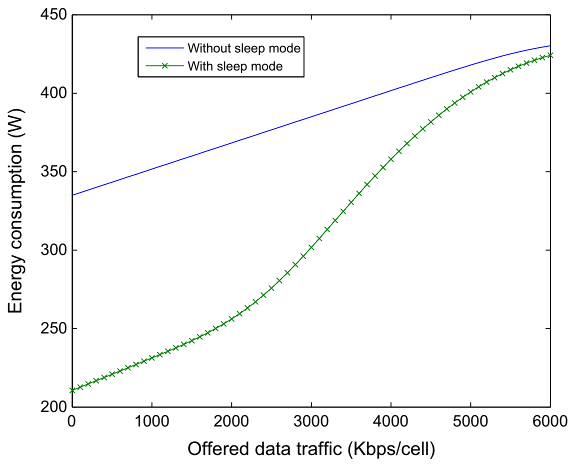

Fig. 4. Energy consumption of an HSDPA base station for different offered data traffic.

4.2. Numerical analysis: MDP resolution

In this section, we study the behavior of the optimal controller and demonstrate its performance on HSDPA networks. We obtain the optimal controller numerically using MDP by calculating the transition probabilities, costs and rewards; value iteration is then used to derive the optimal policy satisfying the constraints. We account for dynamics of users. More details are provided in Section 3. We consider a cell equipped with 4 carriers, each one of maximal capacity equal to 3 Mbps. We limit the maximum number of active users equal to 20 and file size *F* as 2 Mbps for this illustration. Minimum user satisfaction rate is taken as 1.4 Mbps. A target throughput of 1.4 Mbps is sought and QoS is measured as the proportion of users that have a throughput higher than this target.

First, we illustrate the optimal policy developed in Section 3. We show the action chosen by the optimal policy as a function of the number of users in the cell (Fig. 5) for various traffic volumes. We observe that behavior of optimal policy is quite intuitive. When system's state changes from 2 active users, then best optimal policy is to activate only one more carrier for low traffic volume *i.e.*, 2 Mbps and the remaining two carriers are kept switched off to save energy while ensuring user-perceived QoS. Whereas, for high traffic volume *i.e.*, 8 Mbps, the optimal policy is to activate all the carriers in order to achieve required QoS.

Next, we show the impact of target throughput as a function of minimum user rate on the optimal policy by fixing the traffic volume as 2 Mbps (Fig. 6). We observe that implementing sleep mode is the best option for lower target throughput, while for higher targeted throughput, implementing sleep mode is not rational.

Then, by varying the capacity of the cell, we come across different optimal policies for the same system states (Fig. 7). For example, by increasing the cell capacity from 3 Mbps to 4 Mbps, we can switch off more resources for the same number of active users. Thus, the optimal policy does depend not only on traffic volume but also on capacity of the cell.

4.3. Simulations

The analysis of the previous section was based on theoretical models of the radio interface. In this section, we show simulation results by using the network simulator Odyssee [21] in order to assess the performance of the controller in a realistic setting. Odyssee is a large-scale simulator that considers real geographical and site information and implements realistic Radio Resource Management schemes (admission and congestion control, handover, etc.), based on correlated snapshots of 1 s each. Odyssee begins by dropping mobiles in the simulated area, following geographical traffic information. It then associates mobiles to base stations and applies the RRM schemes like power control, interference management and so on, like classical Monte Carlo static simulators and computes QoS measures. However, unlike classical static simulators, it does not draw another independent snapshot, but generates correlated snapshots on the basis of 1 s time interval. It specifically removes users that have finished their service during the past second, applies mobility to existing users and adds new users following a spatial Poisson process. This process is repeated for a long time (thousands of seconds) and the overall performance measures are computed at the end of the simulation by

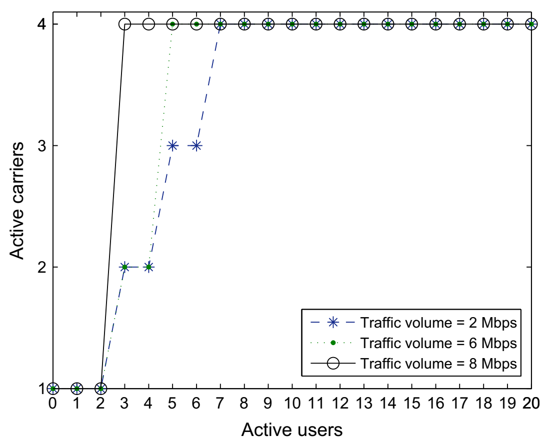

Fig. 5. Optimal policy for different traffic volumes.

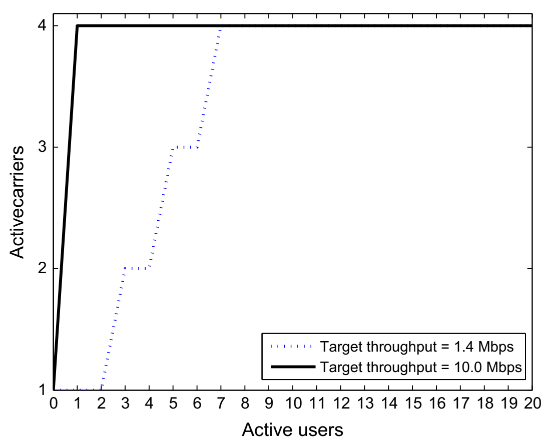

Fig. 6. Impact of target throughput on the optimal policy.

Please cite this article in press as: R. Combes et al., Optimal online control for sleep mode in green base stations, Comput. Netw. (2015), http://dx.doi.org/10.1016/j.comnet.2014.10.031

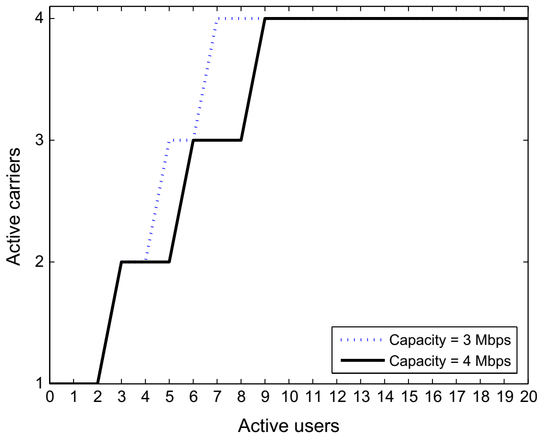

Fig. 7. Optimal policy by varying cell capacity.

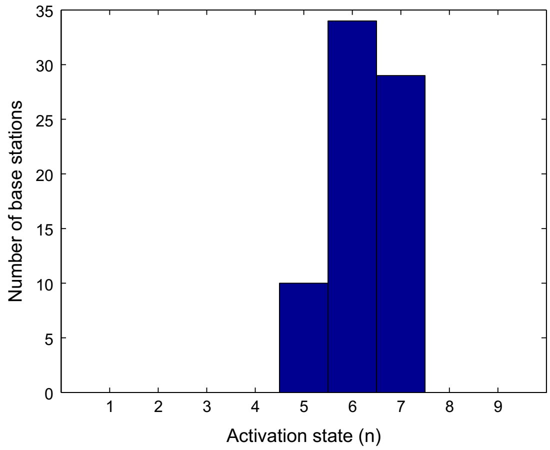

Fig. 9. Histogram of the optimal activation state over the network.

averaging over all the snapshots. A detailed description of Odyssee can be found in [21].

Simulations are carried out on a part of the 2G Paris network carrying voice traffic on both 900 MHz and 1800 MHz frequency bands. Base stations are tri-sectored except one omni-sectored micro base station. Fig. 8 shows the zone chosen for the simulations (i.e., highlighted in the dark rectangle). We implement the sleep mode controller in each base station and simulate the network long enough to reach the steady-state.

Fig. 9 shows the distribution of the optimal transmitter activation state over the base stations of the network. Three different policies are observed over the network, depending on the traffic loads and the base station configurations. As for the optimized hysteresis time, Fig. 10 shows that it has a large dispersion around 8 s. This is due to the heterogeneous nature of the network, as the different base stations have different traffic loads and different configurations (numbers of transmitters).

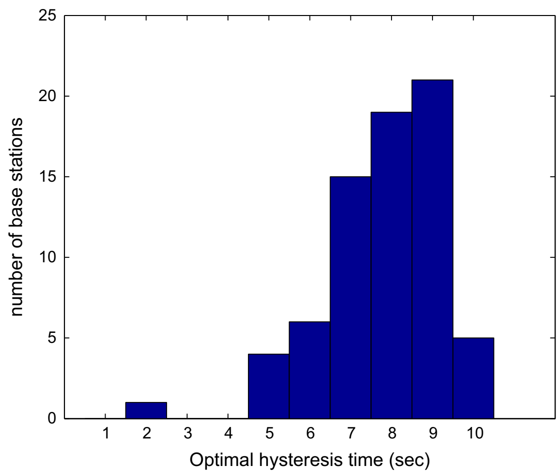

Fig. 10. Histogram of the optimal hysteresis time over the network.

5. Learning the optimal policy

5.1. An ϵ-greedy algorithm

The results presented in the previous section show interesting properties of the controller:

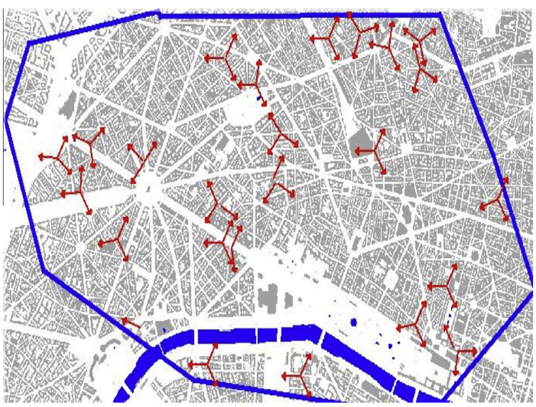

Fig. 8. Simulated area.

1. The optimal policy is a threshold policy on the number of users in the cell, as illustrated in Figs. 5–7. This restricts largely the number of possible policies as only threshold policies are of interest.
2. The threshold value for activating a carrier depends largely on the system parameters (traffic, radio conditions, etc.) as can be shown from the numerical examples (Figs. 5–7) and the simulation results (Figs. 9 and 10). This indicates that we need an online controller that does not rely on an a priori knowledge of the traffic intensity or the radio conditions.

Based on these observations, we propose to use a simple learning algorithm, based on the ϵ-greedy strategy in order to implement online controllers in the base stations. ϵ-greedy strategies are a well known type of policies to solve the exploration vs. exploitation dilemma by choosing the next policy based on previous sequence of plays and rewards (see for instance [14] for an analysis in the case

of independent rewards). Let $\check{\mathbf{S}}$ be the set of all possible threshold policies. The algorithm proceeds as follows:

- Time is divided into discrete steps of equal length during which the sleep mode policy is kept constant. The duration of one step is to chosen long enough to take into account the traffic dynamics (arrivals and departures) and small enough for a fast convergence of the algorithm. We choose in this work an interval duration of 5 min that corresponds to the duration over which base stations usually average the observed traffic and report it to the network.
- At a given time step i, a policy $P(i) \in \check{\mathbf{S}}$ is chosen. During the next time interval, a realization of the cost function g_i is observed (an example cost function is that given in Eq. (24)).
- An estimator of the cost function (empirical average) for each of the policies $p \in \check{\mathbf{S}}$ is updated by:

$$\hat{g}_p(i) = \left(\sum_{j \leqslant i} g_j I_{P(j)=p} \right) \frac{1}{t_p(i)}, \tag{32}$$

  where $t_p(i) = \sum_{j \leqslant i} g_j I_{P(j)=p}$ is the number of times policy p has been selected.
- Given a fixed $\epsilon > 0$, the policy for the next step is chosen as:

$$P(i+1) \sim \begin{cases} \arg\max_{p \in \check{\mathbf{S}}} \hat{g}_p(i), & \text{with probability } (1-\epsilon), \\ \mathcal{U}, & \text{with probability } \epsilon, \end{cases} \tag{33}$$

  where $\mathcal{U}$ denotes the uniform distribution on $\check{\mathbf{S}}$.

It is noted that using standard results on the ϵ-greedy policy in MDPs (see for instance [27]), one can show that for any $\epsilon > 0$, the algorithm converges to the optimal policy. By convergence we mean that when the number of time steps goes to infinity, the proportion of time steps at which the optimal policy is chosen converges to 1. It is noted that the problem at hand is *not* a Multi-armed bandit as studied in [14], since the rewards obtained at successive trials are *correlated*, so that one cannot apply the results of [14] to prove that the number of time steps a suboptimal decision is taken is logarithmic. Despite this fact, we will show that the algorithm works very well in practice.

The algorithm described above and illustrated in pseudo code 1 is valid when traffic is stationary. For slowly changing traffic, we use a straightforward modification of it by calculating the empirical averages $\hat{g}$ on a time window of fixed size τ. Namely at time i, decisions are taken only based on observations during time interval $[i-\tau, i]$.

Algorithm 1. sleep mode learning algorithm

At each time step and for each sector:
- set a sleep mode policy $P(i) \in \check{\mathbf{S}}$ and apply it,
- collect, during all the time interval, the cost function g_i,
- update the estimator of the cost function for each of the policies $p \in \check{\mathbf{S}}$ as in Eq. (32),
- choose the policy for the next step as in Eq. (33).

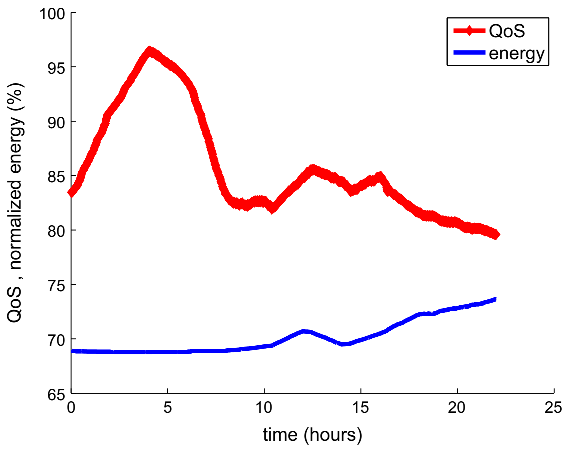

Fig. 11. Performance of the proposed online learning scheme.

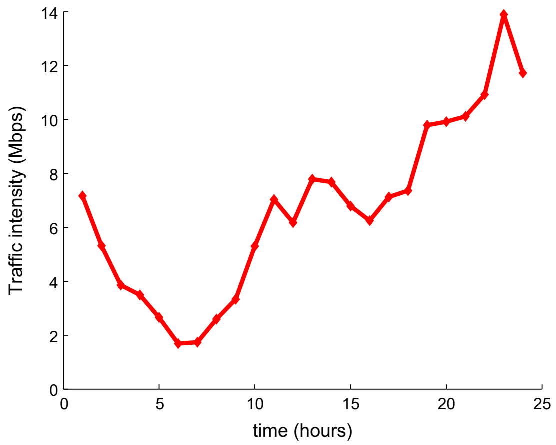

Fig. 12. Traffic intensity as a function of time.

5.2. Numerical illustration of the online controller

In Fig. 11, we illustrate the performance of the proposed algorithm, and represent the QoS and the amount of the total energy used (in percent). As in a real network, the traffic intensity changes over the course of the day. The traffic intensity as a function of time is taken from real network traces and is given by Fig. 12: there is little traffic during the night, and a peak of traffic during the evening. We use the following parameters: $\tau = 1$ hour and $\epsilon = 0.1$. The parameter γ of the cost function is chosen so that at all times, a QoS of 80% or more is guaranteed.

We observe from Fig. 11 that the QoS achieved by the proposed algorithm is indeed always above 80%, and that the amount of energy consumed is an increasing function of the traffic intensity. Namely, during the night less than 70% of the total energy is used, while 75% is used to cope with the traffic peak of the evening. Therefore the proposed algorithm indeed is able to adapt to the daily traffic patterns in an on-line fashion. Furthermore, over the

course of a typical day, the algorithm enables saving more than 25% of the energy needed to run the network.

6. Conclusion

In this paper, we investigated network sleep mode schemes in mobile networks. Based on the observation that base stations have a preponderant part in the overall energy consumption of the network, we focused on switching OFF resources of the cell whenever possible, i.e., if the QoS is not degraded. We derived the optimal sleep mode policy that gives, for each traffic scenario, the best actions to take in each state of the network. We first proposed obtaining the optimal controller by solving an MDP corresponding to the sleep mode policies. We also studied ping pong effect that may appear at the frontier between two capacity regions. We showed how to derive the optimal policy using Markov decision theory. We illustrated the proposed schemes using numerical analysis based on theoretical models of the radio interface, and then demonstrated, by realistic network simulations, how the controllers behave in a heterogeneous network. In order to accommodate large-scale network and diverse traffic configurations, we proposed an online controller that explores the possible policies using a simple ϵ-greedy algorithm and show that it learns the optimal policy rapidly through numerical experiments.

Appendix A. Notation used

A list of the main quantities used in this article is found below.

Notation	Meaning
R	Number of resources
$\mathbf{s}_t$	State as a function of time
$\mathcal{S}$	State space
$\tilde{\mathcal{S}}$	Modified state space taking into account the number of activated resources
$\rho_j(\mathbf{s}\|R)$	load of resource j in state $\mathbf{s}$
c_{DC}	DC to RF power conversion factor
P_{max}	Radio frequency output power of the power amplifier
P_{TRX}	Fixed power consumption of the radio transceiver
P_{cst}	Fixed power consumption due to transport and processing units
F_r	Frontier in the state space associated to resources r
$\mathbf{P}$	Resource activation policy
$\mathbf{P}^\star$	Optimal resource activation policy
$\mathcal{P}$	Set of admissible policies
$\mathbf{Q}$	Transition matrix of the original, non constrained process
$\tilde{\mathbf{Q}}(\mathbf{P})$	Transition of the process controlled by policy $\mathbf{P}$
$q(\mathbf{s}, \mathbf{s}')$	Transition probability between states $\mathbf{s}$ and $\mathbf{s}'$
$\tilde{\pi}$	Stationary distribution of the controlled process
$E(\mathbf{s})$	Energy consumption of the system in state $\mathbf{s}$
$f(\mathbf{P})$	Expected QoS under policy $\mathbf{P}$
n	Number of active users
n_c	Number of active users with radio condition c
b_c	Power consumed by a user with radio condition c
N_{trx}	Number of time slots
N_{sig}	Number of time slots reserved for signaling
P_{max}	Total power transmitted on a carrier
P_{com}	Power used by signaling channels
T	Target throughput
D_c	Throughput achievable by a user in radio condition c
λ_c	Arrival rate of calls for users in radio condition c
μ	Service rate of calls (for any radio condition)
ν	Uniformization rate
F	Expected flow size
$\mathcal{A}$	Set of admissible actions
$g(\mathbf{s}, a)$	Cost incurred by selecting action a in state $\mathbf{s}$
$I_{.}$	Indicator function
T_{off}	Hysteresis time
$\mathcal{U}$	Uniform distribution

References

[1] 3GPP TR 36.814 v0.4.1, Further Advancements for E-UTRA, Physical Layer Aspects, February 2009.
[2] F. Cucchietti, G. Griffa, L. Radice, Eco-efficiency indicator: an operators energy performance indicator, IEEE Telecommunications Energy Conference, September 2007.
[3] L. Sartori, S-E. Elayoubi, B. Fourestié, On the cooperation between 3G and 2G systems, in: IEEE VTC-fall 2007, Baltimore, October 2007.
[4] Q. Song, A. Jamalipour, Network Selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques, IEEE Wireless Commun. 12 (3) (2005) 42–48.
[5] S. Alouf, E. Altman, A.P. Azad, M/G/1 queue with repeated inhomogeneous vacations applied to ieee 802.16e power saving, in: ACM Sigmetrics, 2008.
[6] S.R. Yang and Y.B. Lin, Modeling UMTS discontinuous reception mechanism, in: IEEE Transactions on Wireless Communications, vol. 4, January 2005.
[7] I. Haratcherev, M. Fiorito, C. Balageas, Low-power sleep mode and out-of-band wake-up for indoor Access Points, in: IEEE GreenComm09, December 2009.
[8] H. Claussen, I. Ashraf, L.T.W. Ho, Dynamic idle mode procedures for femtocells, Bell Labs Tech. J. 15 (2) (2010) 95–116.
[9] G. Micallef, P. Mogensen, H.O. Scheck, Cell size breathing and possibilities to introduce cell sleep mode, in: European Wireless Conference 2010, April 2010, pp. 111–115.
[10] J. Kelif, M. Coupechoux, F. Marache, Limiting power transmission of green cellular networks: impact on coverage and capacity, in: IEEE ICC, 2010.
[11] J. Palicot, Cognitive radio: an enabling technology for the green radio communications concept, in: IWCMC, June 2009.
[12] L. Saker, S-E. Elayoubi, H.O. Scheck, System selection and sleep mode for energy saving in cooperative 2G/3G networks, in: IEEE VTC-fall 2009, Anchorage, September 2009.
[13] M.A. Marsan, L. Chiaraviglio, D. Ciullo, M. Meo, Optimal energy savings in cellular access networks, in: IEEE ICC 2009, June 2009.

Please cite this article in press as: R. Combes et al., Optimal online control for sleep mode in green base stations, Comput. Netw. (2015), http://dx.doi.org/10.1016/j.comnet.2014.10.031

[14] P. Auer, N. Cesa-Bianchi, P. Fischer, Finite-time analysis of the multiarmed bandit problem, Machine Learn. 47 (2002) 235256.
[15] S-E. Elayoubi, L. Saker, T. Chahed, Optimal Control for Base Station Sleep Mode in Energy Efficient Radio Access Networks, IEEE Infocom, Shanghai, 2011.
[16] M. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley-Interscience, 1994.
[17] S-E. Elayoubi, T. Chahed, G. Hébuterne, Mobility-aware admission control schemes in the downlink of third generation wireless systems, IEEE Trans. Vehicular Technol. (2007)
[18] T. Bonald, A. Proutière, Wireless downlink data channels: user performance and cell dimensioning, in: ACM Mobicom 2003, 2003.
[19] X. Guo, O. Hernndez-Lerma, Continuous-Time Markov Decision Processes: Theory and Applications, Springer, 2009.
[20] N. Enderlé, X. Lagrange, User satisfaction models and scheduling algorithms for packet-switched services in UMTS, in: IEEE VTC, 2003.
[21] A. Samhat, Z. Altman, M. Francisco, B. Fourestié, Semi-dynamic simulator for large-scale heterogeneous wireless networks, Int. J. Mobile Netw. Design Innovation 1 (3/4) (2006) 269–278.
[22] J. Lorincz, A. Capone, D. Begusic, Impact of service rates and base station switching granularity on energy consumption of cellular networks, EURASIP J. Wireless Commun. Netw. 342 (2012) 1–44.
[23] S. Baek, J.J. Song, B.D. Choi, Performance analysis of push-to-talk over IEEE 802.16e with sleep mode and idle mode, Telecommun. Syst. 47 (3–4) (2011) 291–302.
[24] S. De Vuyst, K. De Turck, D. Fiems, S. Wittevrongel, H. Bruneel, Delay versus energy consumption of the IEEE 802.16e sleep-mode mechanism, IEEE Trans. Wireless Commun. 8 (11) (2009) 5383–5387.
[25] L. Kong, G. Wong, D. Tsang, Performance study and system optimization on sleep mode operation in IEEE 802.16e, IEEE T. Wirel. Commun. 8 (9) (2009) 4518–4528.
[26] L. Saker, S. Elayoubi, R. Combes, T. Chahed, Optimal control of wake up mechanisms of femtocells in heterogeneous networks, IEEE J. Select. Areas Commun. 30 (3) (2012) 664–672.
[27] R. Sutton, R. Barto, Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, 1998.

Richard Combes was born in France in 1985. He received the Engineering Degree from ENST (Ecole Nationale Supérieure des Télécommunications, Paris, France) in 2008 and the Master Degree in Mathematics from the University of Paris VII (France) in 2009. He was a visiting scientist in INRIA in 2012 and a post doc at KTH in 2013. He is now an assistant professor at Supelec, France. His current research interests include wireless networks, game theory and queuing theory. He was the recipient for the best paper award at the Conference on Network and Service Management (CNSM) in 2011.

Salah Eddine Elayoubi received his M.S. in telecommunications and networking from the National Polytechnic Institute at Toulouse, France, in 2001, and his Ph.D. and Habilitation in Computer science from the University of Paris VI in 2004 and 2009, respectively. Since then, he has been working in Orange Labs, the research and development division of France Telecom group. He is now a senior research engineer, involved in several EU-funded research projects. His research interests include radio resource management, modeling and performance evaluation of mobile networks.

Arshad Ali received a B.Sc. in Mathematics and Physics from Punjab University, Lahore in 1997. In 2003, he earned a M.Sc. degree in Computer Science from Punjab University, Lahore. Next in 2009, he received a Master diploma in Information Technology with speciality in Mobile Networks. He completed his Ph.D. degree in Information Technology, Telecommunications and Electronics jointly with Institute of Telecom SudParis and UPMC (Paris VI) in November 2012. Between 2005 and 2008, he served as an Audit Officer in the office of Auditor General of Pakistan. After Ph.D., he worked on energy efficiency for LTE networks in the radio engineering for mobile networks team of Orange labs till November, 2013. Since December 2013, he is working as Assistant Professor at University of Lahore, Lahore, Pakistan.

Louai Saker was born in Latakia, Syria, in 1985. He received his Engineering degree in Electronics and Telecommunications from the Lebanese University, Faculty of Engineering, Tripoli, Lebanon, in 2007 and his Masters degree in Networking at UPMC (Paris 6) and Telecom ParisTech (ENST), Paris, France, in 2008. He received his Ph.D. degree in Electronics and Telecommunications from Telecom SudParis (INT) in 2011 after three years of researching work at Orange Labs. His research interests during the Ph.D. were in the areas of Green Networking where we focus on reducing the energy consumption of mobile networks. Since 2012, he works with Alcatel-lucent in the LTE radio engineering team.

Tijani Chahed holds BS and MS degrees in Electrical and Electronics Engineering from Bilkent University, Turkey, and Ph.D. and Habilitation a Diriger des Recherches (HDR) degrees in Computer Science from the University of Versailles and the University of Paris 6, France, respectively. He is currently a Professor in the Telecommunication Networks and Services department in Telecom SudParis, France. His research interests are in the area of quality of service and teletraffic engineering, both in the wireline and wireless networks.

Please cite this article in press as: R. Combes et al., Optimal online control for sleep mode in green base stations, Comput. Netw. (2015), http://dx.doi.org/10.1016/j.comnet.2014.10.031